

# 1

## *Sociolinguistics as Language Variation and Change*

*Not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity.* (Weinreich, Labov, and Herzog 1968: 188)

In this chapter I introduce fundamental concepts and key constructs of the study of Variationist Sociolinguistics that will be detailed in later chapters. Why approach the study of language from this perspective? What can be learned from this method that cannot be learned from other sociolinguistic methods? A major component of this approach to language is that it is linguistic, but also social and statistical. Why is a combined socioquantitative method useful and desirable?

### **Sociolinguistics**

Sociolinguistics has its roots in dialectology, historical linguistics, and language contact with considerable influence from sociology and psychology (Koerner 1991: 65). This is why it has evolved into an exceptionally broad field. An all-encompassing definition would be that the domain of inquiry of sociolinguistics is the interaction between language, culture, and society. Depending on the focus, virtually any study of language implicates a social connection because without this human component language itself would not exist. However, the scope of sociolinguistics in this expansive interpretation is vast. Sociolinguistics has as many different facets as its roots. Some areas of the discipline put more emphasis on one area (culture); some disciplines put more emphasis on another (education). There is no one sociolinguistics other than the overarching unity of language in use. Depending on which aspect of language in use comes to the fore, sociolinguistics diverges into innumerable subdisciplines.

Every day we speak and write and use a complex, structured system to communicate but at the same time that system is evolving. The fundamental LVC (Language Variation and

Change) question is, *How does this happen?* Weinreich, Labov, and Herzog (1968: 100–101) answered the question by saying, “the key to a rational conception of language change is the possibility of describing orderly differentiation.” This order, yet differentiation, as the normal state of affairs (Labov 1982: 17), the idea that variation is an inherent part of language (Labov 1969: 728), is the foundational maxim of the LVC approach. Differentiation, anomalies, and nonstandard features are easy to spot. In fact, just about everyone likes to talk about the wacky, weird, and/or reprehensible bits of language.

The normal condition of the speech community is a heterogeneous one ... Moreover this heterogeneity is an integral part of the linguistic economy of the community, necessary to satisfy the linguistic demands of every-day life. (Weinreich, Labov, and Herzog 1968: 17)

Variability [is] not.. a nuisance but is a universal and functional design feature of language. (Foulkes, 2006)

Variation in language is most readily observed in the vernacular of everyday life. For example, a teenager says: “that were like sick”; an elderly man recounting a story to his granddaughter says: “you was always workin’ in them days.” Are these utterances mistakes? Are they slang? Are they instances of dialect? An LVC-oriented sociolinguist views such instances of language in use as an indication of the variable but rule-governed behavior typical of all natural speech varieties. The vernacular was first defined as “the style in which the minimum attention is given to the monitoring of speech” (Labov 1972c: 208). Later discussions affirmed that the ideal target of sociolinguistic investigation is “everyday speech” (Sankoff 1974, 1980b: 54), “real language in use” (Milroy 1992: 66). Variation in language can be observed just about everywhere from a conversation you overhear on the street to a story you read in the newspaper. Sociolinguists notice such variations too. In undertaking sustained analysis, what they discover is that people will use one form and then another for more or less the same meaning all the time the language varies. The harder part is to find the order, or the system, in the variation chaos. The way LVC undertakes this is by means of the “linguistic variable.” A linguistic variable is the alternation of forms, or “layering” of forms, in language. A basic definition is “two or more ways of saying the same thing.” A more nuanced, early, definition also mentions that linguistic variables should be structural and “integrated into a larger system of functioning units” (Labov 1972: 8).

Linguistic variables in a given speech community, whether morphosyntactic, phonological, lexical, or discursive, do not vary haphazardly, but systematically. Because it is systematic, this behavior can be quantitatively modeled (Labov 1963, 1969). Analyses of heterogeneous structures within the speech community rest on the assumption that whenever a choice exists among two (or more) alternatives in the course of linguistic performance, and where that choice may have been influenced by any number of factors, then it is appropriate to invoke statistical techniques (Sankoff 1988a: 2). The statistical tools used in the study of variation will be discussed in Chapter 5.

The combination of methods employed in Variationist Sociolinguistics forms part of the “descriptive-interpretative” strand of modern linguistic research (Sankoff 1988a: 142–143). Large-scale studies of variation in speech communities from New York to Norwich have produced extensive bodies of data. The descriptive component requires detailed, critical observation of variation and change. The patterns that have emerged from these undertakings have demonstrated that linguistic change is not only the result

of universal principles but is also shaped by the social context in which it occurs (Labov 1963: 74). This is where the interpretive component of LVC has proven critical. Descriptions of variation can only be understood in context. While sociolinguistic principles prevail wherever you go, each situation provides a unique interpretation. In the case studies in Chapters 7–11 I will demonstrate how the study of different types of linguistic variables must take into account historical, contemporary, and social facts to explain language use.

## The Linguistic Variable

LVC research begins with the observation that language is inherently variable. Speakers make choices when they speak and they alternate among these choices. Take, for example, the use of forms which strikes the ear as nonstandard, unusual, dialectal, or new, as in the examples in Example 1.1.<sup>1</sup>

### Example 1.1

- (a) And then next *mornin'* [In] they were all brought back again. (YRK/002)
- (b) Our car was *like* seven miles from where the entrance was. (TOR/021)
- (c) There was a supply boat *Ø* came down to our cottage everyday. (TOR/036)
- (d) He was like *so* funny and *so* nice. (TOR/054)

These features can only be fully understood if they are examined alongside the relatively unremarkable alternates with which they vary, as in Example 1.2.

### Example 1.2

- (a) And I started work on an *evening* [ɪ]. (YRK/012)
- (b) We were oh probably *about* six miles from it. (TOR/054)
- (c) The people *that* did it were brainwashed. (TOR/069)
- (d) She's *really* funny, and I think she's *really* pretty too. (TOR/021)

Some variables may even have three or more alternates, as in Example 1.3.

### Example 1.3

- (a) I can't remember what that *building* [in] is called. (TOR/008)
- (b) I was on vacation for *approximately* six weeks. (TOR/038)
- (c) I'm only exposed to the people *who* speak the same way that I do. (TOR/016)
- (d) He's *very* funny; he's *very* generous. (TOR/023)

In other words, speakers may vary among various pronunciations of “ing” at the end of words. They may signal approximation with *like* or *about* or *approximately*. They may choose among relative pronouns *that* or *who* or leave it out entirely. They may select *so* or *really* or *very* to intensify an adjective. These choices are potential “linguistic variables.”

**NOTE** Linguistic variables are typically referred to by inserting the phoneme or morpheme or word that is variable inside parenthesis, i.e. variable (ing), (ly), (that), (so), etc. Phonetic realizations are represented inside square brackets, e.g. [n]. Phonemes are represented inside forward slashes, e.g. /n/.

A linguistic variable in its most basic definition is two or more ways of saying the same thing. An important question is, What does it mean to say two things mean the same thing? One time a student asked this question: what is the difference between a synonym and a linguistic variable? Let us explore this distinction. Synonyms are different lexemes with the same referential meaning as in Example 1.4:

#### Example 1.4

- (a) *car, automobile, vehicle, wheels*
- (b) *girl, lass, chick, sheila, babe, doll, skirt*

A more restrictive definition of synonymy would require that two synonyms are completely interchangeable in every possible context. In reality, most are not. For example, *lass* is primarily used in Scotland and northern England, *chick* is used in North America, *sheila* in Australia, whereas *girl* is not confined to a particular variety of English. For many practical purposes, such as with the production of dictionaries, it is customary to adopt a looser kind of definition for synonym. Near synonyms are lexemes that share an essential part of their sense, as in Example 1.5:

#### Example 1.5

- (a) *interesting, intriguing, fascinating, absorbing, spellbinding, engrossing*
- (b) *striking, arresting, unusual, out of the ordinary, remarkable, salient*

But this is not the whole story. Linguistic variables must also be alternatives (i.e. options) within the same grammatical system which have the same referential value (meaning) in running discourse (Sankoff 1988a: 142–143). Although some variants may differ subtly in meaning and distribution, if they are part of a linguistic variable they will be members of a structured set in the grammar. Moreover, the choice of one variant or the other must vary in a systematic way – this is what is meant by structured heterogeneity. There is difference, but there is structure to it. Different ways of saying more or less the same thing may occur at every level of grammar in a language, in every variety of a language, in every style, dialect, and register of a language, in every speaker, often even in the same discourse in the same sentence. In fact, variation is everywhere, all the time. This is why it is referred to as “inherent” variation (Labov 1969: 728). Now, consider a more in-depth definition of the linguistic variable:

- two different ways of saying the same thing;
- an abstraction;

- made up of variants;
- comprising a linguistically defined set of some type:
  - a phoneme
  - a lexical item
  - a structural category
  - a natural class of units
  - a syntactic relationship
  - the permutation or placement of items
- although its delineation can be at any level of the grammar, the variants of the variable must have a structurally defined relationship in the grammar;
- they must also co-vary, correlating with patterns of social and/or linguistic phenomena.

Synonyms could be a linguistic variable. However a linguistic variable is more than simply a synonym. Deciding which forms co-vary meaningfully in language is actually a lot trickier than you would think.

### Mini Quiz 1.1

- Q1 How would a variationist sociolinguist explain the following example?  
 “There *was* two of us. Yeah, that’s right there *were* two of us.”
- (a) Alternation in styles.
  - (b) Free variation.
  - (c) Linguistic variation.
  - (d) Random differentiation.
  - (e) Bad grammar.
- Q2 Which of the following provides an example of two variants of a linguistic variable?
- (a) And we said, “if you join the club, you must go to church.”
  - (b) He’d light a furnace for to wash the clothes.
  - (c) He was awful homesick, you know, my Uncle Jim.
  - (d) To prove I could do it, I had to prove that I could do it.
  - (e) There’s two girls on my street who have pink hats.

To this point this discussion has focused on the technical description of the linguistic variable. However, there is an entirely different side to linguistic variation that does not come from the mechanics of the linguistic system but involves issues of stigma and salience that come from the external evaluation of language by its users – us humans. There is no reason for a velar sound to be superior to an alveolar sound. There is no reason for a synthetic construction to be better than an analytic one. There is no inherently terrible thing about a double negative. However, there is an absolutely insidious view that certain ways of saying things are better than others. This comes down to the social interpretation of language use.

Most people are convinced that linguistic features are good or bad. For example, here is Sara Kempt, aged 49, in Toronto, Canada (c. 2003), in Example 1.6.

## Example 1.6

... and I think the natural inclination of anybody is to get lazy and sloppy and not think. So I th— there's more and more slang, and people dropping their Gs and things like that, just that ... frankly grates on me. I hate it! *Then again, I find myself doing it sometimes.* (TOR/027)

Another fascinating thing about linguistic variables is that people are often completely unaware that they use them, particularly when certain of the variants are not part of the standard language. For example, this is Gabrielle Prusskin, aged 55, in Toronto, Canada (c. 2003). The interviewer has just asked her what she thinks about the word *like*, as in Example 1.7.

## Example 1.7

It's usually young females um when every other word is "like" and it drives me insane. I just *like* I hate it. (TOR/054)

**TIP** One way to find a linguistic variable is to look for the words that occur most frequently in data. Are there other ways of saying the same thing? If language is always in flux, then it is just a question of finding out what is on the move in a particular place and time.

Linguistic variables inevitably involve variants that have social meaning. These are typically called "sociolinguistic variables." Sociolinguistic variables are those which can be correlated with "some nonlinguistic variable of the social context: of the speaker, the addressee, the audience, the setting, etc." (Labov 1972c: 237). One variant might have overt social stigma, e.g. "I *ain't* got it", another might entail authority, e.g. "You *must* listen", or prestige, e.g. "I *shall* tell you a story." Yet another variant may be neutral, e.g. "I *have* it." These social evaluations may differ markedly from one community to the next, from one country to the next, from one variety to the next, from one social situation to the next. It may even be the case that one person's admired pronunciation will be another person's loathed one. The patterns of a linguistic variable in the speech community tell the story of how the speech community evaluates the variants of the variable and in so doing this reveals how society is organized and structured. Which groups talk to each other? Which groups do not? How a linguistic feature is socially evaluated often has to do with its history as well. Which groups have been in the community a long time? Which groups are new? Language use is a reflection of the society in which it is embedded and the time period in which it occurs.

**NOTE** One time I went to a conference in the United States with my then current group of British graduate students. One of them had a strong accent from a variety of somewhat modest prestige in the United Kingdom. She was shocked to be told, repeatedly, how lovely her accent was. Similarly, I was chagrined to discover that my own middle-class Canadian accent – unremarkable in Canada – was heard as an entirely unbecoming American accent in the United Kingdom.

The primary empirical task of Variationist Sociolinguistics is to correlate linguistic variation as the dependent variable with independent variables (Chambers 2003: 17). The dependent variables are the features of the linguistic system that vary (e.g. the varying pronunciations of the same phoneme, the choice of relative pronoun, the selection of an intensifying adverb). Independent variables are the features associated with the variation. They can be external to the grammar, out in the world, relating to aspects of the social context, situation, community setting, or register. They can also be internal to the grammar, relating to the linguistic environment such as the grammatical category of the word, the type of subject in the clause, or its function.

Patterns in language are observed using a two-part undertaking: (1) find socially and linguistically significant factors that impact variation, and (2) correlate them with general social forces (Labov 1972c: 42). The patterns that arise are used by the analyst to interpret and explain the phenomenon under investigation. The fact that linguistic differentiation in communities has been consistent for different linguistic features and that these patterns repeat themselves across different situations in time and space have given rise to a series of “classic” sociolinguistic patterns from which Variationist Sociolinguistic inquiry has sprung. These patterns provided a baseline for all subsequent research and have informed several new generations of research-based study.

The study of sociolinguistics as LVC is unique in sociolinguistics in two ways: (1) its overriding goal and (2) its methodology. LVC research attempts to solve one of the great paradoxes of language in use – the fact that language is always changing.

The basic LVC procedure is the following:

- Observation – hear and/or see variation in language use;
- Identification – select the linguistic variable for study;
- Reconnaissance – determine if the variation occurs and where;
- Systematic Exploratory Observation:
  - What is the inventory of forms?
  - What are the patterns?
  - When does the variation occur and under what circumstances?
  - Who uses the variation and how?
- Test hypotheses, claims, and observations;
- Interpret and explain the variable patterns, social and linguistic.

To discover the relevant factors (social and linguistic) which give rise to “speakers or writers’ sustained and repeated exercise of their linguistic facilities in producing large numbers of sentences” (Sankoff 1988a), the data are analyzed using statistical modeling. This method enables the analyst to ask and answer the following questions:

- Which factors are statistically significant (i.e. not due to chance)?
- What is the relative contribution of the independent factors tested in the model, i.e. which factor group is most significant or least?
- What is the order (from more to less) of factors within the independent factors (predictors), the constraints or constraint hierarchy?
- Does this order reflect the direction predicted by one or the other of the hypotheses being tested?

Notice that Variationist Sociolinguistics has an essentially multiplex nature: on the one hand, empirical and data-based; on the other hand, scientific methods and statistical testing; but there is a third component. Linguistic patterns can only be understood through interpretation. Explanation in sociolinguistics can only happen when statistics are used in conjunction with a strong interpretive component, grounded in real-world language use.

### Mini Quiz 1.2

- Q1 The primary empirical task of Variationist Sociolinguistics is to:
- (a) define linguistic variables
  - (b) relate linguistic variables with each other
  - (c) correlate independent variables with each other
  - (d) correlate linguistic variation as the dependent variable with independent variables
  - (e) correlate linguistic variation as the independent variable with dependent variables in society.

## Linguistic Change

If I talk to, say, my grandfather, like I talk to one of my friends, he'd just be like, "what?" (TOR/023)

One of the driving forces of Variationist Sociolinguistics is the search for general principles that govern linguistic change. If one form appears to be replacing the other, either in time or along some economic, demographic, or geographic dimension (Sankoff and Thibault 1981: 213) then this may be an indication of language change in progress. Consider the way people talk about the weather. It is often the case that it is either cold or hot outside. When the temperature is extreme in one direction or another people will typically intensify their descriptions. For example, if it is particularly cold a person might say, "It's very cold today!" But would a young person say it the same way? Probably not. A younger person (at least in Canada in the early twenty-first century) is more likely to say: "It's so cold today!" In contrast a middle-aged person is more likely to say: "It's really cold!" If these observations can be substantiated across a wide number of people (e.g. Tagliamonte 2008b) (see Chapter 9), this may be evidence for ongoing evolution of a subsystem of grammar – generational change. This is why linguistic data from different age groups in the same speech community, or different communities in the same country, or even communities in different countries in different circumstances, provide important evidence for understanding how language change may be happening.

All languages change through time. We do not really know why this is, but it is a characteristic of all human languages. They also change in different ways in different places. (Trudgill 2003: 7)

Linguistic change typically proceeds in "an ordered set of shifts in the frequency of application of the rule in each environment" (Labov 1982: 75). What this means is that the rate of use of a particular form, e.g. *very*, *really* or *so*, is not the most important observation. Instead, the contexts in which these forms occur – their patterns of use – is the key element



in tapping into linguistic change. For example, *very* and *really* can occur as attributive or predicative intensifiers, as in Example 1.8a, but *so* can only occur as predicative, at least in current prescriptive accounts of English.

### Example 1.8

- (a) It was a *really* hot day [attributive] and like on the way there I started to feel *really*, *really* weak [predicative]. (TOR/011)
- (b) I was *so* hungry [predicative]. (TOR/013)

These patterns of use are the fundamental units by which linguistic change occur (Labov 1982: 75) (see Chapter 3). Moreover, if the relationship of these environmental contexts can be captured this provides a critical measure for comparison – referred to as a “constraint” on variation. In this case the constraint is the difference between predicative vs. attributive position. Similarities and differences in the significance, strength, and ordering of constraints (the constraint ranking) offer a microscopic view of the grammar from which we can infer the structure (and possible interaction) of different grammars. The various statistical techniques in the LVC toolkit enable the analyst to assess and evaluate the competing influences and in so doing interpret the path of development of language through time and space and social structure (see Chapter 5).

**TIP** What is the difference between a “variable” and a “factor” or “factor group”? In sociolinguistics “variable” is reserved for the “linguistic variable”, the feature that varies and that is under investigation (i.e. the dependent variable). “Factors” are the aspects of the social or linguistic context that influence the variable phenomena (i.e. the independent variables). In statistics these are referred to as “predictors.” To avoid confusion I will continue to use the field’s current standard term “factor group” or “factor.”

LVC analysis asks the question, How can a variable linguistic phenomenon be explained? In this type of analysis the critical component is that the data come from the recurrent choices speakers make in the course of production. In this way, each choice is viewed not simply as an instance or token of use, but as a choice made within the context of the grammar from which it comes. When a large body of repeating tokens is part of the analysis, the choices can be assessed statistically so as to uncover the meaningful patterns of use (Cedergren and Sankoff 1974; Labov 1969; Poplack and Tagliamonte 2001: 89). The choices are taken to represent the (underlying) variable grammar of the speaker as well as the grammar of the speech community to which she belongs (Poplack and Tagliamonte 2001: 94). The goal to investigate language use in the context of language structure is what makes an LVC analysis “accountable.”

## The Principle of Accountability

A foundational concept in the Variationist Sociolinguistic approach and one that sets it apart from other methods is the “principle of accountability” (Labov 1966: 49; 1969:737–738, fn. 20; 1972c: 72). This is where the analysis begins. Say the analyst is interested in the use of

the relative pronoun *who*. The principle of accountability dictates that in addition to examining *who* itself, the analyst must also take into account all the other potential variants within the relative pronoun system. Accountability requires that all the relevant forms in the subsystem of grammar that you have targeted for investigation, not simply the variant of interest, are included in the analysis. The idea is that the analyst cannot gain access to how a variant functions in the grammar without considering it in the context of the subsystem of which it is a part. Then, each use of the variant under investigation can be reported as a proportion of the total number of relevant constructions, i.e. the total number of times the function (i.e. the same meaning) occurred in the data (Wolfram 1993: 206).

### **Circumscribing the Variable Context**

The focus of investigation in LVC is the place in the linguistic system that is variable; in the case of the relative pronoun *who*, it would be the relative pronoun system. However, delimiting a subsystem of grammar is often not an easy task. In many cases, there will be more than one variant (e.g. relative *who*, *that*, etc.). Often one of the variants will be zero which makes spotting these variants particularly difficult. Inevitably, there will be contexts that are ambiguous and cases where the same form can have an entirely different meaning (e.g. in addition to being a relative pronoun, *that* can also be a complementizer, a locative, an expletive pronoun, etc.).

According to the principle of accountability, it is necessary to circumscribe the data to only those contexts that are functionally parallel as well as variable. The task is to determine, sometimes by a lengthy process of trial and error, which tokens are in and which are out.

The final decision as to what to count is actually the final solution to the problem at hand. (Labov 1969: 728)

Perhaps the trickiest problem is to determine which forms mean the same thing. The LVC approach to form/function asymmetry is that distinctions in referential value or grammatical function among different surface forms are often neutralized in discourse (Sankoff 1988a: 153).

Many “functions” can be carried out by several different “forms” and the question of who, when and why become immediately pertinent in accounting for those actually used. (Sankoff 1988a: 151)

Determining where the linguistic variable varies is called “circumscribing the variable context” (Poplack and Tagliamonte 1989a: 60) or “the envelope of variation” (Milroy and Gordon 2003: 180). This task requires that the analyst identify the total pool of utterances in which the feature varies. In dealing with form/function asymmetry Variationist Sociolinguistics becomes a descriptive-interpretive research enterprise (Sankoff 1988a: 149). In other words, a variationist study has two ways of looking at data: (1) the distribution of forms and (2) the identification of the linguistic function of each form.

Contexts that do not vary but are categorically encoded with one or other variant are not included in the analysis of variation. These are the “don’t count” cases (see Blake 1994). This does not mean the categorical contexts are not important. Knowing which areas of the

grammar are categorical and which are variable is a critical part of interpreting patterns and explaining linguistic phenomena (see Labov 1972b: 815; Smith 2001); however, categorical contexts cannot be part of an analysis of variation. That would be mixing apples with oranges.

These procedures accentuate that the study of variation is not interested in individual occurrences of linguistic features, but requires systematic study of the recurrent choices an individual makes (Poplack and Tagliamonte 2001: 89). Analysis of these recurrent choices enables the LVC analyst to find the underlying patterns that lead to one choice or the other. A “pattern” is representative. It refers to “a series of parallel occurrences (established according to structural and/or functional criteria) occurring at a non-negligible rate in a corpus of language use” (Poplack and Meechan 1998b: 129).

The analysis of data can only be as good as the data provided by the extraction process.  
(Wolfram 1993: 203)

In sum, the LVC approach does not simply study the features of language that are attention-grabbing or unusual all by themselves. It also studies their alternates. This is the real challenge: is there variation? And if so, where exactly does it occur? Where is the variable? Labov characterizes this process as a “long series of exploratory manoeuvres” (Labov 1969: 728–729). The case studies in Chapters 7–11 will provide examples of these comprehensive methodological practices.

Along with the key construct of the linguistic variable and the key maxim of the principle of accountability – the next foundational pillar of the LVC approach to sociolinguistics that differentiates it from other methods in sociolinguistics is its quantitative method. Once an adequate number of choices has been taken into account, the patterns of use have been identified and coded, then the analyst can discover the system in the variation using statistical modeling (Cedergren and Sankoff 1974; Labov 1969).

### Mini Quiz 1.3

Q1 What is the difference between an occurrence and a pattern?

- (a) A pattern cannot be discerned by systematic and exhaustive quantitative analysis of variability.
- (b) A pattern occurs at a negligible rate in a corpus of language.
- (c) A pattern is a series of heterogeneous occurrences.
- (d) A pattern is representative.
- (e) A pattern cannot reveal the grammatical provenance of forms.

### *Frequency*

How often does a linguistic form occur? The rate of occurrence of a feature is an important first step in understanding variation. The frequency of a feature is dependent on the contexts that are included in the calculation. It is not sufficient to embark upon the

counting enterprise by simply counting the number of times a variant of interest occurred in a body of data. Why? First, how often a feature occurs depends on the quantity and nature of the data. Second, the number of occurrences of an item in one body of materials cannot be compared to the number of occurrences in a different body of materials unless there is a way to “normalize” the two data sets. One way to do this is by counting the total number of words in each data set. However, the problem is that data sets can differ markedly in terms of their contents, making overall counts problematic for comparison. For example, if you wanted to study the future in English and you had two interviews of the same individual, but one conversation had focused on upcoming events and the other on past memories you would have a very different tally of future forms per number of words. If you counted how many times a particular future variant occurred out of the total number of times the person used future temporal reference in each data set, proportions of one variant or another would likely be pretty much the same. This is why the frequency of a feature is determined by counting how many times it occurred as a proportion of the number of places where it could have occurred – in other words, a distributional analysis.

The starting point of a variation analysis is a survey of the overall distribution of forms with the same function. The first question to ask is, How many of the variant of interest occur? The second, Out of how many? In other words, How often was there a *possibility* that the variant would occur in the data under investigation? Only with this knowledge can the number of instances of the variant in question become interpretable.

The next step is to determine the independent factors that may influence where the variant under investigation can occur.

### Mini Quiz 1.4

Consider the data below and examine the relative pronouns.

- (a) And there was a man named Mr Pape who used to come around with fruit and vegetables to the door. (TOR/035)
- (b) There was a man that played the piano. (TOR/035)
- (c) ... she could tell you people who lived here like forty years ago like. (TOR/099)
- (d) Then I found out that I had torn the muscle. (TOR/025)
- (e) There was a man with a horse and cart that used to deliver them water from the lake for washing. (TOR/035)
- (f) You can't really trust a man who doesn't pay child support. (TOR/070)
- (g) But there was a man that sat in the middle of the streetcar in a little sort of box. (TOR/049)
- (h) There was also a man came up-and-down the street with a small cart and a bell and he sharpened scissors. (TOR/049)

Q1 How many variants are there?

Q2 Which sentence doesn't belong?

Q3 What is the proportion of *who*?

### Constraints

The fundamental unit of change is not the rule but the environmental constraint within the rule.  
(Labov 1982: 75)

Variationist Sociolinguistics views the behavior of the dependent variable as it distributes across a series of cross-cutting factors, whether external (social) or internal (grammatical). To gain access to this information, it is necessary to determine how the choice of a particular variant is influenced by different aspects of the contexts in which it occurs (Sankoff 1988b: 985).

The frequency of variants will fluctuate considerably from one individual to the next or one situation to the next. A less educated person might use more *got* for stative possession, *You got a pen?* than someone with a college degree, *Have you got a pen?* A person in a conversation with a friend might use it more than with a boss. My children might use it more than I. A person from England might even use a different construction: *Have you a pen?* However, the patterns of variation within the same speech community remain stable (Sankoff 1988a: 153). If *got* is used at all, it tends to be used with second-person subjects. Thus, frequency and pattern are distinct measures of a variable phenomenon. Frequency tends to fluctuate due to external conditions but the tendencies in grammatical patterns stay the same. This finding has been reported and repeated in the literature from the earliest days of quantitative methodology to the most recent. For example, in a study of bare English origin nouns in Spanish, Torres-Cacoullos (2003: 323) concluded “even with typologically similar language, variable rule analysis can reveal details of the grammar that constitute conflict sites, even when rates for variants are similar.” Meyerhoff and Walker (2007), studying the varieties of English spoken in Bequia, a small island in the Caribbean, reports that

raw frequencies of vernacular variants may fluctuate, but language-internal constraints persist. Among the many implications these findings have for the study of linguistic variation and change, perhaps the most crucial is that it affirms “the validity of modeling variable rules in a community grammar, rather than as an aggregation of idiolectal norms.” (Meyerhoff and Walker 2007: 346)

What does it mean to count? The sociolinguist tallies each instance of all the variants of the variable then to a consideration of the constraints. The sociolinguist examines how often each variant of the variable occurs in each context. This type of analysis is called a comparison of marginals. It shows the percentages of the different variants in the data (an overall distribution of forms and a factor by factor analysis); for example, in the relative pronoun data above, *who* occurred 3 times out of 7. The overall distribution of *who* is 3/7 or 42.8%. The distribution of relative pronouns could also be examined in terms of animacy of the antecedent. How often does *that* occurs with animate antecedents, *a girl that danced*, as compared to inanimate ones, e.g. *a bar that opened up*. This information would provide a view of how the dependant variable (the choice of relative pronoun) is influence by the animacy of the antecedent (a factor by factor analysis).

### Practice LVC

Consider the excerpt in Example 1.9 from an interview with Carla Brennan, age 19, born in Toronto, Canada, c. 1991.

## Example 1.9

[107] ... but this one- one girl, Sophie, she had like posters, you know. You put stuff- pictures up in your locker. It's like of a guy or *something*[ŋ], right? She had elephants up in her locker, like ... oh, this is so funny! 'Kay, she had all these elephants. Elephants are cute. She had buttons and elephant shirts, *everything*[ŋ]. Pure elephants, right. And I go like, "Grow up!" I don't know, she just took it too far, okay? Anyway, uh one day we were *looking*[n] at *National Geographic* and we saw, uhm, I don't know, uh, elephants like *being*[n] whatever, stabbed and uh, I don't know what they were *doing*[n], just *using*[n] their bodies and all this stuff and my friend starts *ripping*[n] out the pages. She's *laughing*[n], right. And she's *laughing*[n]. I go, "What are you *doing*[n]?" And then her ... this is Sophie's locker right cross an- and then she starts *shoving*[n] the pictures in. I go, "Andrea, don't, don't!" She's like *laughing*[n] her head off. She put them in and then ... this is like before lunch, like the lunch hour. So, we waited and watched Sophie come to her locker, open it up and the pictures fall out. She looks at them and she goes, "Oh, uh!" She's just so. I don't know. She's so like that, just "uh!" And she looks at her friends and she goes, "Who did this?" And then I'm like ... *trying*[n] not to laugh, right? And we're just *watching*[n] this whole *thing*[ŋ]. And we're like ... And Andrea's *going*[n], "Ha, ha ha!" And I go, "Don't laugh, don't laugh." And she's like, "Who did this? This isn't very funny guys." And then she goes and throws them in the garbage. And her friends are like, "We didn't do it Sophia. I don't know what you're *talking*[n] about." But they were *laughing*[n] too, right? It's terrible (laughing) ....

In this excerpt Carla alternates between the velar nasal with [ŋ] and the alveolar nasal with [n] in words ending in "ing." The variants of (ing) have been highlighted.

To begin an LVC analysis, start with the principle of accountability. Where does the [ŋ] variant occur and where could it have occurred, but did not, i.e. [n] occurred instead? To do this count the realization of the sounds for each word where the ending "ing" occurs. In this way each of the variants – in this case [ŋ] and [n] – may be examined according to the proportion it represents of the total number of contexts. What is the overall distribution of [n]? It is 15/18, or 83%.

An LVC analysis proceeds by circumscribing the context of variation. The first step is to recognize and remove categorical tokens from subsequent analysis. Note, for example, that there is one token that will never vary between [n] and [ŋ], but only ever occurs as [ŋ]. It is the lexical item, *thing*. Why is there no variation? Because if *thing* were pronounced as *thin* it would be a different word entirely. Once a categorical context has been identified, remove it. In this case, remove all tokens of the word *thing*. This requires an adjustment of the calculation of variability, namely 14/17. The proportion changes slightly to 82.3%. Had we continued the analysis to all of Carla's data we would have inevitably found many tokens of the word "thing." If these were included in the analysis of variable (ing) they would skew the data toward [ŋ] since none of the tokens would occur as [n]. The second step is to identify contexts that cannot be unambiguously identified as one variant or the other. In spoken data, not all variants that qualify as potentially variable can be reliably assessed because they are inaudible, neutralized, or there is some other cause of indeterminacy. As it happens, another token must be removed, the lexical item *being*. When speakers pronounce this word there is too much reduction to distinguish either [n] or [ŋ] for the segment to be assessed. Thus, tokens of *being* are also set aside. The adjusted variation now has an overall distribution of 13/16. The proportion changes to 81.2%. As the analyst hones the data like this which forms are in and which are out becomes an important series of steps in the study.

A critical aspect of circumscribing the variable context is to consistently document the procedures that were undertaken in extracting the data and deciding which contexts are

included in the analysis. This is the only way that the study can be confidently replicated by the next researcher. The whole comparative sociolinguistic enterprise rests on the consistency of the analyses (see Chapter 6).

An LVC analysis tests for the constraints on variation, whether social or linguistic. As it happens, there is a strong linguistic pattern that underlies the variation between [ŋ] and [n]. Notice that the only tokens which contain the velar variant are the lexical items “something” and “everything”, both indefinite pronouns. In contrast, all the other words ending in “ing” are verbs, particularly progressive verbs. I will return to consider this observation further in Chapter 7.

In sum, an LVC approach to sociolinguistics is based on data that is taken to represent the speech community and uses methods that enable the analyst to discover the structure of the variable in the grammar. When you embark on an analysis, you never know what may turn up in the data.

### Mini Quiz 1.5

- Q1 Take another look at the Carla excerpt. At least one other linguistic variable can be identified. Which of the following is it?
- (a) Variation between *be like* and *go as* quotative verbs.
  - (b) Frequent use of discourse marker *like*.
  - (c) The tag *right*.
  - (d) Use of relative *that* for collective nouns (e.g. There’s this recycle club that hangs out).
  - (e) All of the above.

**TIP** People often ask me how I find linguistic variables. If you listen and you look carefully you will find them. Watch for tips throughout the book.

## Evolution of the Linguistic Variable

The definition of the linguistic variable in LVC research began with semantic equivalence, i.e. two ways of saying the same thing. At the word level, this definition provides a relatively straightforward case of semantic correspondence since even if pronunciation differs, the word is still the same. However, at other levels of grammar, the axiom of “mean the same thing” becomes more problematic. In morphology, suffixes may be present or not, as with variable (s), e.g. *I say* vs. *I say*s, or variable (ly), e.g. *go slow* vs. *go slow*ly. Moreover, these differences can have extralinguistic as well as linguistic connotations. The zero variant of variable (ly) is often regarded as American. At the same time presence or absence of the suffix is conditioned by nature of the adverb as abstract vs. concrete (e.g. *go slow* vs. *think slowly*) (see further discussion in Chapter 8). Use of *like* is stigmatized and associated with youth, yet it too has structured patterns of use (see Chapter 9). Use of different intensifiers may seem to have distinct meanings, e.g. *very cold* vs. *really cold*, but the perspective of usage in the

speech community reveals that one form is replacing another (see Chapter 11). How can the analyst deal with linguistic variables at all levels of grammar in a systematic manner?

Establishing functional equivalence beyond the level of phonetics-phonology is problematic. Lay people and linguists alike will argue strongly for meaning differences when presented with potential variables, even when they are framed in near identical phrases. Do the two sentences in Example 1.10a–b mean the same thing?

### Example 1.10

- (a) I think she'll *be* cheeky. (YRK/041)
- (b) I think she's *gonna be* pretty cheeky. (YRK/041)

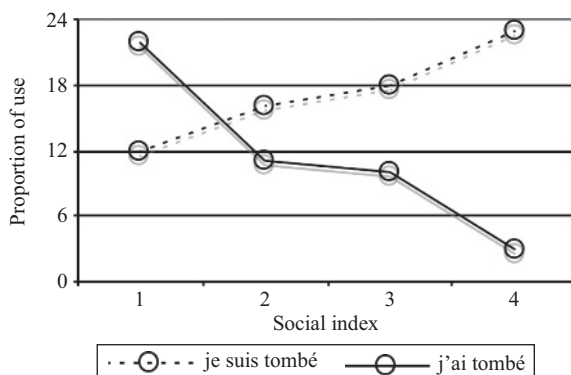
To study this type of variation, the definition of the linguistic variable has come under successive revision. Sankoff (1973) was the vanguard of analyses of variation above the level of phonology. By the early 1980s an extensive debate had arisen (Lavandera 1978; Romaine 1984). Was it appropriate to study syntactic variables or not?

Sankoff and Thibault (1981: 207) laid the groundwork for an accountable approach to this issue by introducing the notion of weak complementarity. This is the idea that linguistic variables can be identified by their distribution across the speech community rather than by the fact that they mean the same thing. They argued that many types of linguistic change do not arise from forms with a common meaning:

Change does not only occur through drift and perturbation of grammatical systems, gradually proceeding from one closely related form to another. It also occurs by the forcible juxtaposition of grammatically very different constructions whose only underlying property in common is their usage for similar discursive functions. It is this relatively violent type of change, which is probably just as prevalent as the gentle diffusion of rule weightings across time and space, which provides so much difficulty for formal grammatical explanation ... (Sankoff and Thibault 1981: 207)

In reality, an LVC analysis begins with the observation that where one variant is used more often another variant is used less. When this observation is made of syntactic, semantic or discourse-pragmatic features, form/function correspondence cannot be sustained because variants involved in the same change may not mean precisely the same thing. However, if they are members of the same structured set in the grammar of the speech community these patterns can be *observed*. The criterion for identifying weak complementarity is a correlation between occurrence rates and some extralinguistic factor of individual speakers such as age, sex, or social index. This is an entirely different measure than is usual in variation analysis, but it can be used as a diagnostic step. It is based on the assumption that certain basic discourse functions will be fulfilled at the same rate from one individual to the next in a given body of data. Take for example the case of auxiliary *avoir* vs. *être* for the verb *tomber*, “to fall” in the Canadian French, *Je suis tombé* vs. *J’ai tombé*. The former is sanctioned by the standard language while the former is nonstandard. Sankoff and Thibault (1981: 211) calculated the rate of use of one variant over the other by line of text and then correlated it with a social index measuring “how important it is to have a mastery of the standard or legitimized variety of the language” on a rising scale of 1–4 (Sankoff and Laberge 1978), as in Figure 1.1.





**Figure 1.1** Rates of *avoir* and *être* usage with “tomber” per thousand lines of transcription.

Source: Sankoff and Thibault 1981: 211, Figure 1.

Figure 1.1 shows that where one form is used less, the other is used more and vice versa – demonstrating weak complementarity of *avoir* and *être*. The alternation shows meaningful social and linguistic trends in this speech community. The analyst can assume that meaning differences between them have been neutralized and can proceed to analysis of variation.

While debate continues with regard whether or not quantitative methods are appropriate for the analysis of certain types of variation, LVC researchers have carried on much in the same spirit as Rickford (1999):

The prospects for ... carrying the “New Wave”<sup>2</sup> into syntax seem promising but not easy. However, there is no reason to limit our goals and methods to those that require the least effort and/or imagination. This is no way to run a revolution. (Rickford 1999: 32)

One of the expanded notions of equivalence is structural equivalence where two forms are studied in a single variable as long as they are found in the same type of context in the language. General extenders (GEs) are a good example. GEs occur at the end of a sentence, in a specific type of syntactic template and are typically used to generalize to a set, as in Example 1.11. Another important diagnostic is to establish that individuals use alternate forms. The variants in Example 1.11a–c come from the same speaker in the same conversation.

### Example 1.11

- (a) A lot of the course centers around immunity and inflammation and stuff like that. (TOR/135)
- (b) I have to focus on my breathing and stuff. (TOR/034)
- (c) Yeah stuff like that like heparin and all that stuff. (TOR/135)

The study of discourse-pragmatic linguistic variables overlaps into conversational analysis and qualitative methods. In this area of research, analysts are primarily focused on the different pragmatic functions of one form or another. Consider the forms *actually*, *really*, and *in fact*, as in Example 1.12 (Waters, in preparation).

## Example 1.12

- (a) [5] Really? [026] Yeah. It was *actually* pretty dry. (TOR/026)  
 (b) [5] You were in high school right? [026] Yeah, I'm *really* pretty much just sick of downtown. (TOR/026)

This type of alternation is at center stage in this controversy. Many researchers believe that these forms mean two entirely different things. In some cases, of course they do, as in Example 1.13a. This example has intensifying function. In other cases there is ambiguity, as in Example 1.13b, where the context could have either intensifying or lexical function. Note that ambiguity is one of the key factors in linguistic change (see Chapter 3).

## Example 1.13

- (a) That's a lot of fun. I *really* enjoyed doing that. (TOR/004)  
 (b) I *really* respected that 'cause most people wouldn't do that. (TOR/004)

The two forms have social value: *really* is casual, vernacular while *actually* is formal, learned. Cathleen Waters recently studied these forms using quantitative methods (Waters, in preparation). Carefully circumscribing the variable context using the principle of accountability and the notion of weak complementary, she demonstrated that in contemporary British and Canadian English *really* is developing and *actually* is fading away. Strong linguistic constraints underlie the use of one or the other: *really* tends to occur in negative contexts and this correlation is becoming stronger in apparent time. This suggests systematic linguistic developments in both of these major varieties of English. Only quantitative methods and diagnostic tests for different processes in linguistic change could have revealed the unique attributes of this system in the grammar of contemporary English.

As LVC research has turned toward linguistic variables that are part of grammaticalizing systems (see Chapter 10), a case can even be made for sometimes including in the analysis contexts that fall *outside* the envelope of variation. Why? Because the phenomenon under investigation may have evolved out of another system. To understand how this has occurred, it becomes necessary to include the broader context of the evolving system in the analysis. Schwenter and Torres-Cacoullos (2008) argue that changing grammatical features make an important contribution to the problem of semantic equivalence. In tense/aspect variables, for example, there is variation across the boundaries of the linguistic variable. Consider the form/function asymmetry of forms used for future temporal reference in English:

- in function, a single form covers a range of meanings  
 e.g. *going to* in English; it is a verb of motion and a future marker
  - He is *going to* the store.
  - It is *going to* rain.
- in form, different forms serve the same grammatical function  
 e.g. *going to* and *will* in English both mark future temporal reference
  - It is *going to* rain.
  - It *will* rain.

This means that the variable context cannot be circumscribed by grammatical function because a single form may cover a range of meanings along a developing pathway. Instead the

variable context must be extended to include not only all the functions of the extant system (whatever the system under investigation happens to be) but also the array of forms that fulfilled the same function at different stages of the change, as well as meanings that are known to have evolved over the course of the grammatical change (see also Aaron 2010).

There will always be nuances of meaning that are conveyed more by one word, form, or sentence than the other, especially when systems of grammar undergo change. These are the blurry margins of language that must be acknowledged and taken into account. Depending on where you place your analytic lens – syntactic structure vs. speaker agency for example – different aspects of language will come into focus. In the end, the onus is on the analyst to provide defensible arguments to demonstrate relevant social and linguistic correlations. As weak complementarity demonstrates, the proof is in the pudding.

**NOTE** What do you think? Analyze my use of *actually* in this book (if the copy-editor leaves them in). Are there any variants? What differentiates the use of *actually* from *really*? Do they really mean different things? Note the following variant from Chapter 9: *only the uses that truly frame constructed dialogue should be included*. Did I really write that? You may also examine the quotes from famous sociolinguists found throughout the book. What do they use?

## The Importance of Accountability

An analyst must know what the counts and the calculation of distribution of counts means so that the information can be interpreted in the appropriate way. As I mentioned earlier in this opening chapter, you cannot simply count the number of times something occurs because this does not tell you very much unless you know how many times that something occurs in the body of material. Corpus linguistics typically uses counts per X number of words, e.g. 10 per thousand, 10 per 10 000 etc. However, in LVC research it is critical to know how a variant is influenced by a particular type of context compared to another. This requires knowing the distribution of a feature (variant) out of the total number of contexts where it could have occurred but did not. This is how correlations are established.

What happens if you don't follow the principle of accountability? An early example of how different ways of calculating frequency led to divergent results comes from the study of variable verbal (s), e.g. *He say Q* vs. *He says*. In a series of narratives told by ex-slaves in the 1940s in the United States, Schneider (1983: 105) reported that the -s variant was used most frequently with third person singular at a rate of 72%. Yet a study of exactly the same data reports that third person singular has a rate of 12% (Brewer 1986: 136). How can this discrepancy be explained? Easily. The two researchers used completely different methods for analysis. Brewer counted only the inflected verbs for each grammatical person and then calculated the proportion that each grammatical person represented out of all -s inflected verbs. In contrast, Schneider employed accountable methods. He calculated the proportion of -s out of all inflected *and* uninflected verbs in each person, i.e. how many third person singular verbs had -s out of all third person singular verbs in the data. The counts are simply not comparable. Only Schneider's calculation provides information that can be used to assess the propensity of a grammatical person (i.e. third person singular) to receive an -s inflection (see Poplack and Tagliamonte 1989b: 54).

**Table 1.1** Count of all quotative types with *be like* as a quotative.

|                                | %   | N   |
|--------------------------------|-----|-----|
| Indirect speech                | 24  | 186 |
| Direct speech                  | 69  | 533 |
| Sound/gesture                  | 2.8 | 22  |
| Hypothetical                   | 3.8 | 30  |
| Writing                        | .25 | 2   |
| Total number of <i>be like</i> |     | 773 |

**Table 1.2** Distribution of *be like* according to type of quotative, i.e. viewed as a proportion of the total of each type.

|                             | %  | N       |
|-----------------------------|----|---------|
| Indirect speech             | 69 | 186/268 |
| Direct speech               | 56 | 533/955 |
| Sound/gesture               | 55 | 22/40   |
| Hypothetical                | 39 | 30/77   |
| Writing                     | 39 | 2/17    |
| Total <i>be like</i> tokens |    | 773     |
| Total quotative contexts    |    | 1357    |

Another example to illustrate the importance of accountability comes from the use of quotative *be like* (see Chapter 9). One of the main factors influencing the choice of *be like* is the nature of the quotation. How do we establish this? If we count the number of times *be like* is used in a data set and then divide the counts of *be like* according to type of quotative in the data set, we get the results in Table 1.1.

This type of count tells you the number of *be like* tokens for each of the different quotative types. However, it does not tap the patterning of *be like* with regard to quotative type because not all of the quotative types are in the count. To determine how quotative type influences the use of *be like* a different type of count is required. You need to count all the quotative types and then count how often *be like* (as opposed to *say*, *think*, *go* or some other quotative) occurred for each type. Table 1.2 shows this result.

The two counts produce diametrically opposed results. Notice that in Table 1.1 it looks as though direct speech is a prime location for the use of *be like*: 69% of *be like* tokens occur with direct speech. However, the explanation for this result is that quotatives that introduce direct speech actually represent the lion's share (70%) of the quotatives in the data, as you can see from the counts in Table 1.2. There are 955 direct quotes in the data out of a total of 1357 quotative frames. However, if we want to know how the type of quotative may influence the selection of *be like* we need to know how often *be like* occurs out of the total number of times each type of quotative occurs. Table 1.2 shows these counts. What you observe is that *be like* is used more with indirect speech than any other type of quotative.

In sum, different ways of counting provide totally different types of information. When counts differ across contexts then simple counts alone will lead analysts to incorrect

interpretations. Comparisons of studies based on different ways of counting will lead to faulty interpretations. It is critical for analysts to explain precisely how their data were analyzed so that discrepancies of this type can be revealed. Furthermore, this is the only way to ensure that findings are comparable.

## Language Variation and Change and Linguistic Theory

Where is variation in a theory of grammar? Formal theory makes a distinction between I-language and E-language. I-language is the internalized knowledge of an individual speaker whereas any conception of language as something external to the speaker is E-language. Formal theories typically ignore variation, linguistic change, and the transmission of social meaning. As far as Chomsky is concerned, E-language “has no place ... in the theory of language” (Chomsky 1986: 26). It would seem then that to implicate variation in a theory of grammar it would be necessary to situate it in I-language. A recent trend in linguistics is to reconsider variation, what has been referred to as “optionality” or as the “probabilistic grammar.” Adger and Smith (2005: 149) put it this way: “given widespread, structured variation, how is the mental grammar (I-language) organized so that variation arises?”

Since the early 1990s further developments intent on marrying LVC and theoretical linguistics have emerged (e.g. Cornips and Corrigan 2005; Henry 1998; Meechan and Foley 1994). This enterprise holds promise for the contention that the two frameworks have complementary functions (e.g. Mufwene 1994). The new term of reference is “sociosyntax” (Cornips and Corrigan 2005: 20). Researchers advocate that collaborative research between LVC practitioners and theoretical linguists could herald a paradigm shift, particularly if such joint ventures could lead to uncovering the “E-language forces that are actuating a change and even ... possible outcomes” (Fasold 2004: 223).

LVC research confirms that variation is inherent in the individual, the group, the community, and beyond. The best demonstration of inherent variability is alternation of variants in the same individual in the same conversation, as you see in many of the examples in this book. Thus, variation is appropriately situated in the idiolect. But what type of choice mechanism is responsible? In some cases, variation is sensitive to regional distinctions, such as might be subsumed in the notion of competing dialect grammars (Kroch 1989, 2003; Kroch and Taylor 1997). Low-level dialect differences can be quite marked, suggesting a model of competing parameters (Henry 1995). Some variation also has robust supra-local patterns. At the same time general complexity and linear processing effects are apparent. When speakers have a choice between alternative forms they choose the variant which minimizes processing complexity (Grondelaers *et al.* 2009; Hawkins 1994; Jaeger 2008). Thus, some of the regularities and tendencies typical of inherent variability may not necessarily be components of I-language (Sankoff 1978b: 251). On the other hand, while lexical choice can be dialect specific (e.g. *lorry* vs. *truck*), contextual effects (e.g. NP vs. personal pronoun) are more likely to be structural. Speaking for phonological variables, Labov argues that “these variables are in the grammar, they are constrained by the grammar, and they cannot be described apart from the grammar” (Labov 2001a: 84). Variable structures at the morphological and syntactic level are often constrained by semantic distinctions and/or structural configurations whose development can be traced in the history of the language. Incremental differences from one dialect to the

next may reflect pathways of change. Thus, some linguistic variables, at least, are heavily embedded in the grammar. Pintzuk (2003: 525) argues that orderly variation in conjunction with the fact that statistical patterns fit formal syntactic models “strongly suggests that a coherent theory relating grammar and usage can and should be formulated.” It would seem that Variationist Sociolinguistics can have important insights into which linguistic specifications are the relevant ones. This is one of the challenges of the next 40 years of LVC research – how to make this happen.

## Exercises

### Exercise 1.1 Spotting linguistic variables

The sentences in Example 1.14 were extracted from my archive of speech data on UK dialects (Tagliamonte 1996–1998, 1999–2001, 2000–2001, 2001–2003):

#### Example 1.14

- (a) I’ve got to cycle all the way back and then this afternoon I’ll be cycling back up again! You have to keep those thoughts to yourself.
- (b) She’s got a coarse tongue in her.
- (c) I’ve a feeling that they don’t know.
- (d) There’s a certain amount of people what just stay but there’s people who’s out for the bank holiday.
- (e) Youngsters gets far too much.
- (f) Them was built for to feed everything.
- (g) I don’t know her; I haven’t seen her.
- (h) Eventually she come back and it all got sorted out.
- (i) I went down to the market last Monday.
- (j) I think it’s gonna get worse before it’ll get better.

**Question 1** List three sentences that show alternation of different forms for the same function. Indicate the sentence letter and each of the forms (i.e. variants) below.

- (a) \_\_\_\_\_
- (b) \_\_\_\_\_
- (c) \_\_\_\_\_

**Question 2** List three obvious nonstandard features.

- (a) \_\_\_\_\_
- (b) \_\_\_\_\_
- (c) \_\_\_\_\_

**Question 3** List three sentences that have only standard features.

**Question 4** List a sentence which has three nonstandard features in the same sentence.

## Exercise 1.2 The variable context

---

The complementizer *that* is highly variable in contemporary English, as in Example 1.15 (see also Chapter 5).

### Example 1.15

My grannie says *that* her da was very low, waiting on him, you know. She says  $\emptyset$  the Sunday nicht afore my granda was drowned, she was over at – up at haim, she calls it. (PVG/007)

In order to study this variable it is first necessary to define the variable context.

The parameters in Tagliamonte and Smith (2005: 298) are the following: “All (apparent) matrix + complement constructions where *that* or *zero* were possible, i.e. where the target matrix verb occupied this canonical slot.” Such tokens are included as “count” cases. Notice, however, that some of the same matrix verbs may appear in other contexts where they function entirely as “parentheticals.” These are constructions that have been bleached “to the point where the phrase acts as an adverbial” because they are “transportable to positions other than that which they could occupy if they were only functioning to introduce a complement” (Thompson and Mulac 1991b: 239–241). Such constructions are not included in the analysis. They are “don’t count” cases.

**Question 1** Examine the data in Example 1.16. Then decide which contexts are in and which are out.

### Example 1.16

[021] I think he met me, actually, yeah. [Interviewer] Right. How did you come across one another? [021] Well, I was sitting in the – I would be about twenty-five, I think, and I was sitting in the Empire Theatre in Maryport, on my own, watching the picture. (CMK/021)

[019] I think she won the Waterloo Cup, some big, prestigious thing. I hae nae much an idea about greyhound-racing, but– We called the old boat Silver-Anna, I think. (MPT/019)

[002] Sure, even the people in the south o this island I think really, if they were getting without being intimidated and all, they would hae went in under the Commonwealth. You-know. I think they would. (PVG/002)

IN:

---

OUT:

---

## Mini Quiz Answers

1.1 Q1: *c*

Q2: *d* complement deletion in first clause; overt complementizer in the second clause.

1.2 Q1: *d*

1.3 Q1: *d*

1.4 Q1: *3*

Q2: *d – it is a complement clause*

Q3: *3/7, 42.8%*

1.5 Q1: *c and a*

## Notes

- 1 All the examples used in this book are identified by unique codes identifying the corpus from which they come, e.g. TOR = Toronto; in some cases the corpus, the individual speaker, a single digit indicator, e.g. “t”, and sometimes the age and sex of the speaker. A listing of all the corpus codes can be found in Appendix A. All names are pseudonyms.
- 2 By “new wave” Rickford is referring to the methodology of Variationist Sociolinguistics which is a pun on the field’s premier conference whose acronym is “NWAV” and stands for “New Ways of Analyzing Variation.”

## Background Reading

Chambers 2003; Fasold 1969, 1971, 1972, 1984; Holmes 1990, 2001; Labov 1963, 1966, 1969; Milroy 1980, 1981, 1984, 1987; Romaine 1994; Trudgill 1972a,b, 1974a,b, 2000; Wardaugh 2002; Wolfram 1969, 1974.